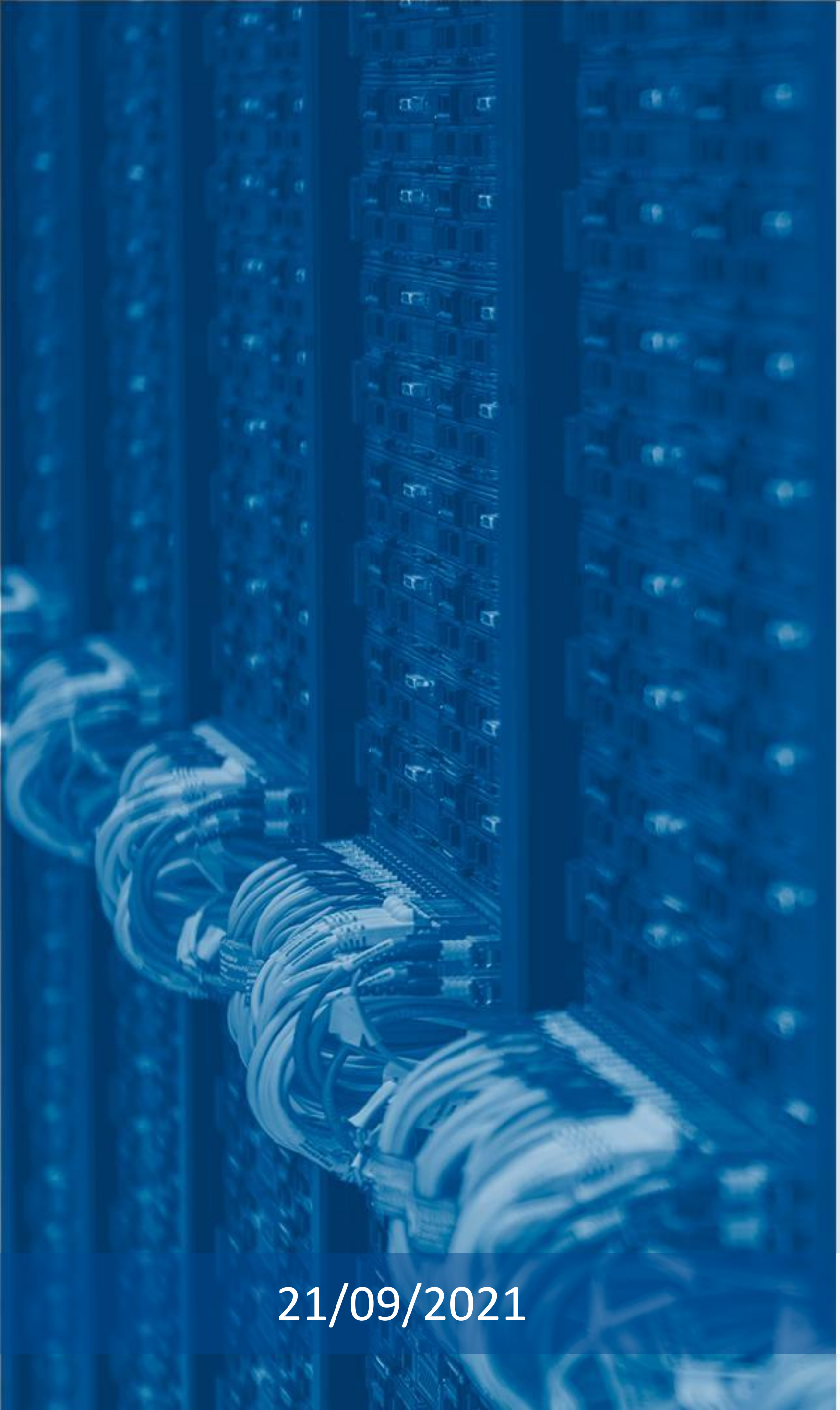


Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Computational efforts to improve the OpenIFS efficiency towards the exascale era

Xavier Yepes-Arbós
Mario C. Acosta

esiwace
CENTRE OF EXCELLENCE IN SIMULATION OF WEATHER AND CLIMATE IN EUROPE

21/09/2021

19th workshop on HPC in meteorology

Introduction

Introduction

- The increase in the forecast capability of Numerical Weather Prediction (NWP) is strongly linked to the spatial resolution to solve more complex problems.
- This requires a large demand of computing power and it might generate a massive volume of model output which implies:
  - Data must be efficiently written into the storage system.
  - No more offline post-processing is affordable due to the size of the “raw” data.
  - A high cost of storage systems due to the huge data size.
- In this context, the improvement of the computational efficiency of NWP models will be mandatory.

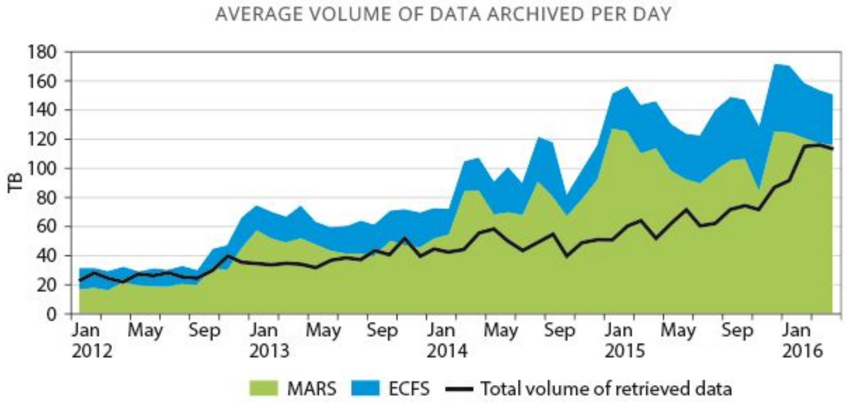


Figure source: ECMWF

OpenIFS overview

- Not much attention was paid on improving I/O of NWP models because it did not use to be an issue, such as in OpenIFS.
- OpenIFS used a sequential I/O scheme, which is not scalable for high resolution grids, and even less, for future exascale machines.
- OpenIFS is derived from the Integrated Forecasting System (IFS), an operational global meteorological forecasting model and data assimilation system developed and maintained by the European Centre for Medium-Range Weather Forecasts (ECMWF). Although OpenIFS has the same forecast capability of IFS, the data assimilation functionality has been removed.

Improvement efforts

In order to improve the efficiency of OpenIFS towards exascale computing, three different efforts are presented, focusing on the efficient I/O management, but also on scalability:

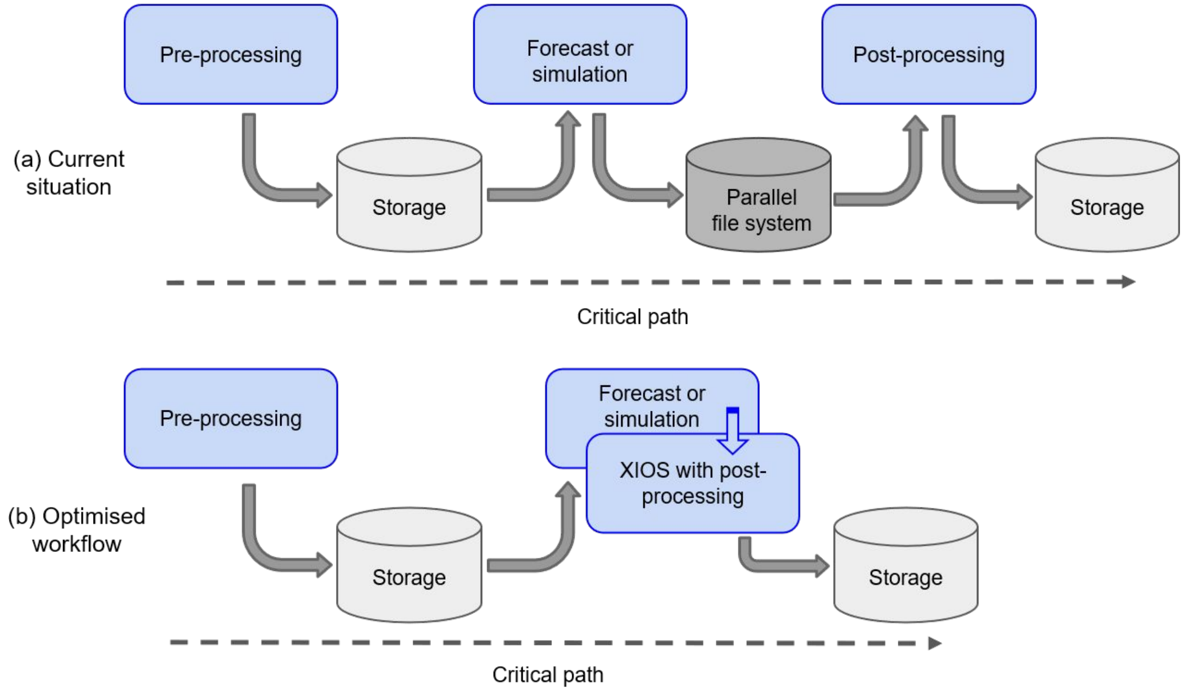
- OpenIFS-XIOS integration.
- Lossy compression.
- Scalability tests.

OpenIFS-XIOS integration

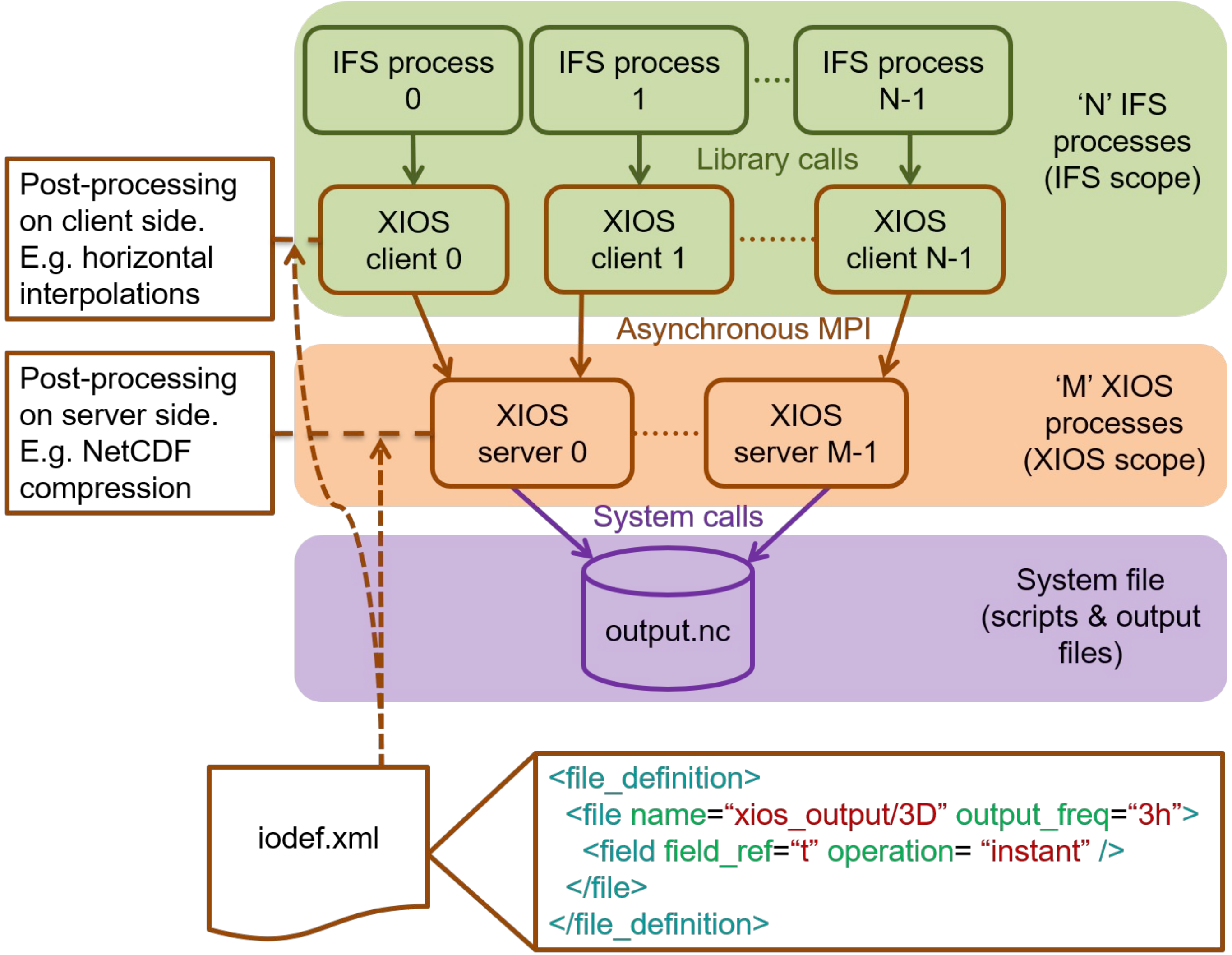
Barcelona
Supercomputing
Center
Centro Nacional de Supercomputación

Objective: Integrate XIOS

- The I/O issue is typically addressed by adopting scalable parallel I/O solutions such as XIOS.
- The XML Input/Output Server (XIOS) is an asynchronous MPI parallel I/O server developed by the Institute Pierre Simon Laplace (IPSL).
- XIOS is a widely I/O tool used in the climate community because of these features:
  - Output files are in netCDF format.
  - Written data is CMIP-compliant (CMORized).
  - It is able to post-process data inline to generate diagnostics.
- XIOS is thought to address:
  - The inefficient legacy read/write process.
  - The unmanageable size of “raw” data by implementing inline post-processing.



OpenIFS-XIOS integration scheme


IFS process 0
IFS process 1
IFS process N-1
'N' IFS processes (IFS scope)
Library calls
XIOS client 0
XIOS client 1
XIOS client N-1
Post-processing on client side. E.g. horizontal interpolations
Asynchronous MPI
XIOS server 0
XIOS server M-1
'M' XIOS processes (XIOS scope)
Post-processing on server side. E.g. NetCDF compression
System calls
output.nc
System file (scripts & output files)
iodef.xml

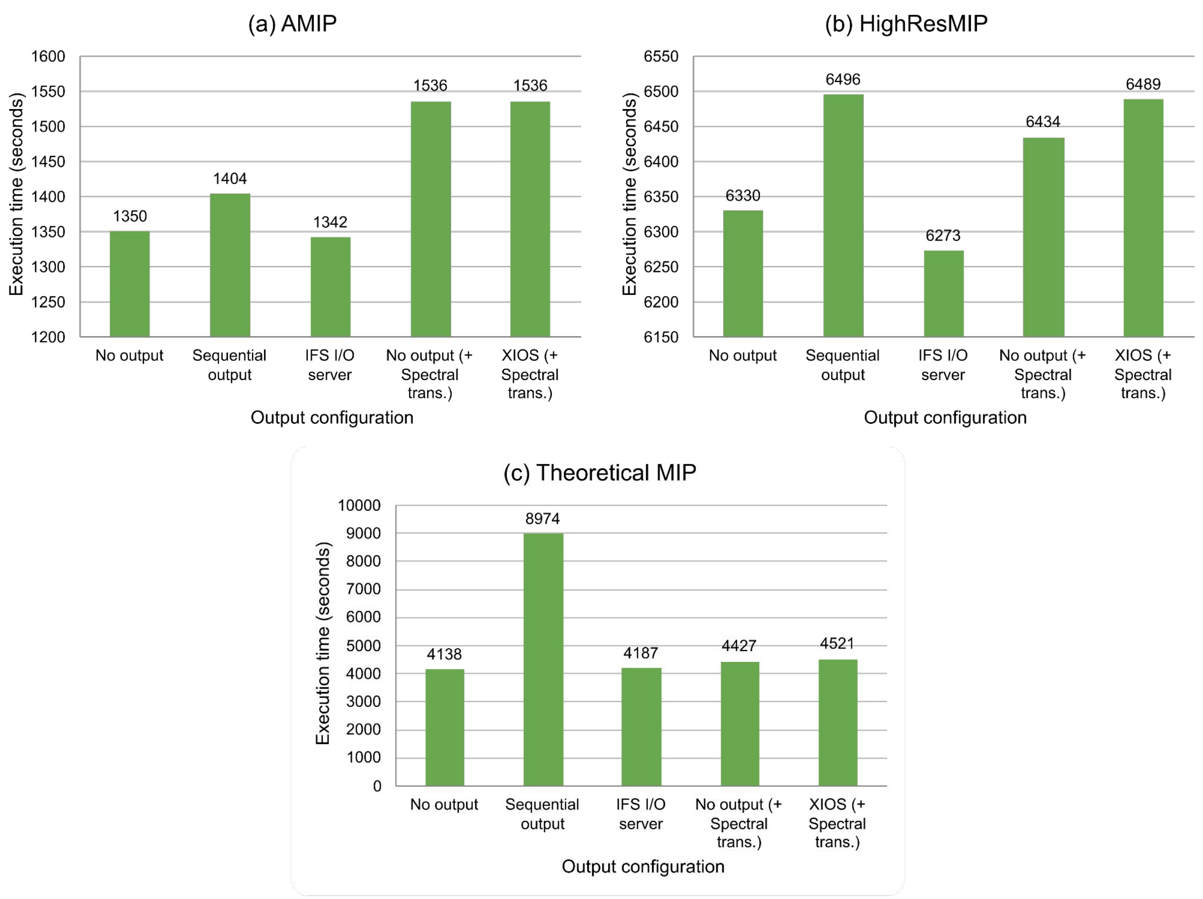
```
<file_definition>
 <file name="xios_output/3D" output_freq="3h">
  <field field_ref="t" operation= "instant" />
 </file>
</file_definition>
```

OpenIFS-XIOS integration summary

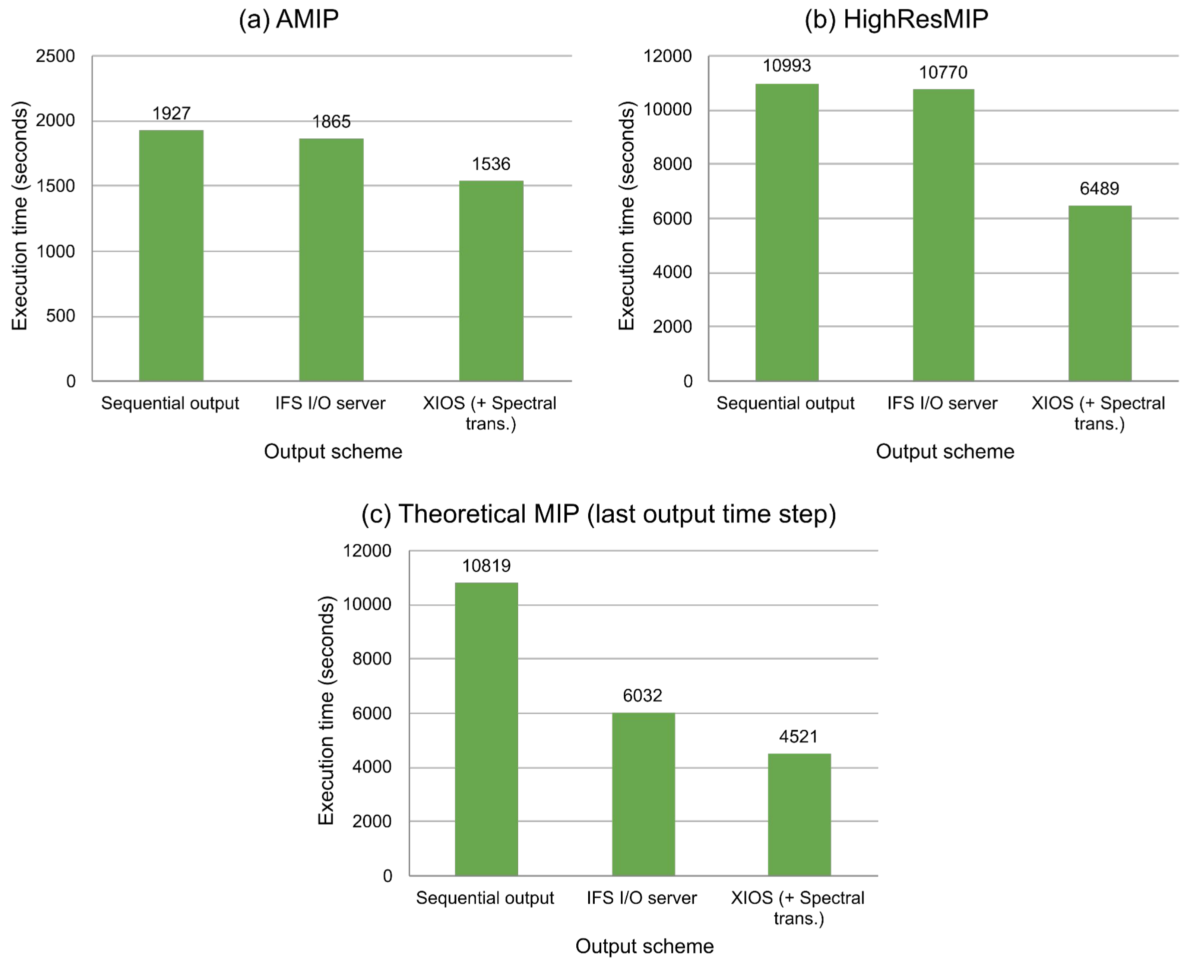
- Scientific highlights:
  - Both grid-point and spectral fields (transformed to grid-point space using TRANS) are supported.
  - All surface and 3D fields can be output.
  - Different vertical levels are available: model, pressure, theta and PV levels.
  - No longer needed to set up the FullPos namelist (NAMFPC).
  - FullPos spectral fitting is available.
  - Physical tendencies and fluxes output (PEXTRA fields) are also supported.
- Computational performance highlights:
  - In-depth benchmarking: the overhead of outputting data through XIOS is really small if using enough computational resources.
  - A profiling and performance analysis was done to detect potential bottlenecks.
  - Two different optimizations are available (switchable in the XIOS XML namelist):
    - Computation and communication overlap.
    - Sends from OpenIFS to XIOS either in double or single precision.

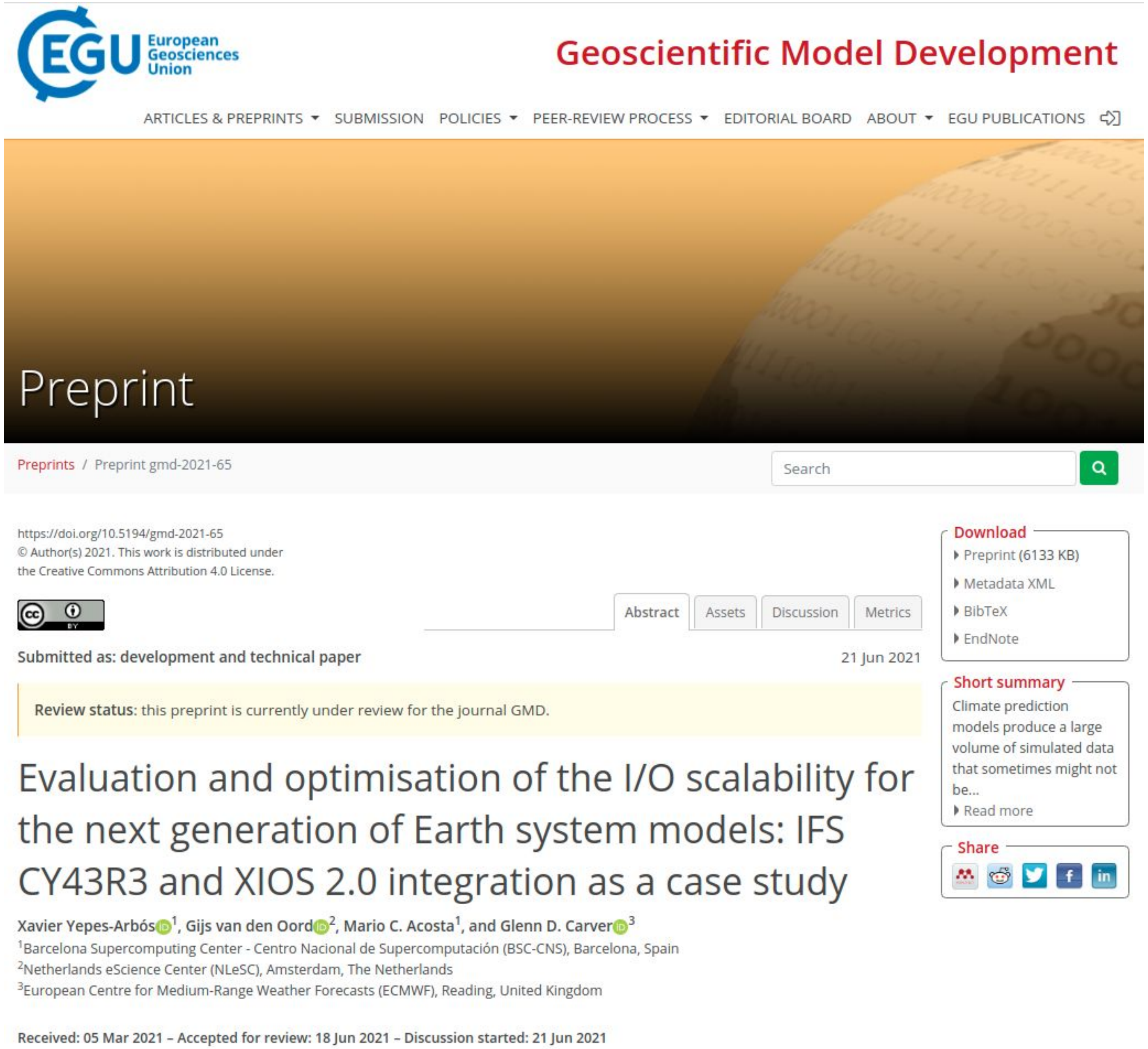
Computational performance of the integration

Output schemes comparison



Output schemes comparison including post-processing





https://doi.org/10.5194/gmd-2021-65

Submitted as: development and technical paper

21 Jun 2021

Review status: this preprint is currently under review for the journal GMD.

Evaluation and optimisation of the I/O scalability for the next generation of Earth system models: IFS CY43R3 and XIOS 2.0 integration as a case study

Xavier Yepes-Arbós[1], Gijs van den Oord[2], Mario C. Acosta[1], and Glenn D. Carver[3]
[1]Barcelona Supercomputing Center - Centro Nacional de Supercomputación (BSC-CNS), Barcelona, Spain
[2]Netherlands eScience Center (NLeSC), Amsterdam, The Netherlands
[3]European Centre for Medium-Range Weather Forecasts (ECMWF), Reading, United Kingdom

Received: 05 Mar 2021 – Accepted for review: 18 Jun 2021 – Discussion started: 21 Jun 2021

Short summary

Climate prediction models produce a large volume of simulated data that sometimes might not be...

Read more

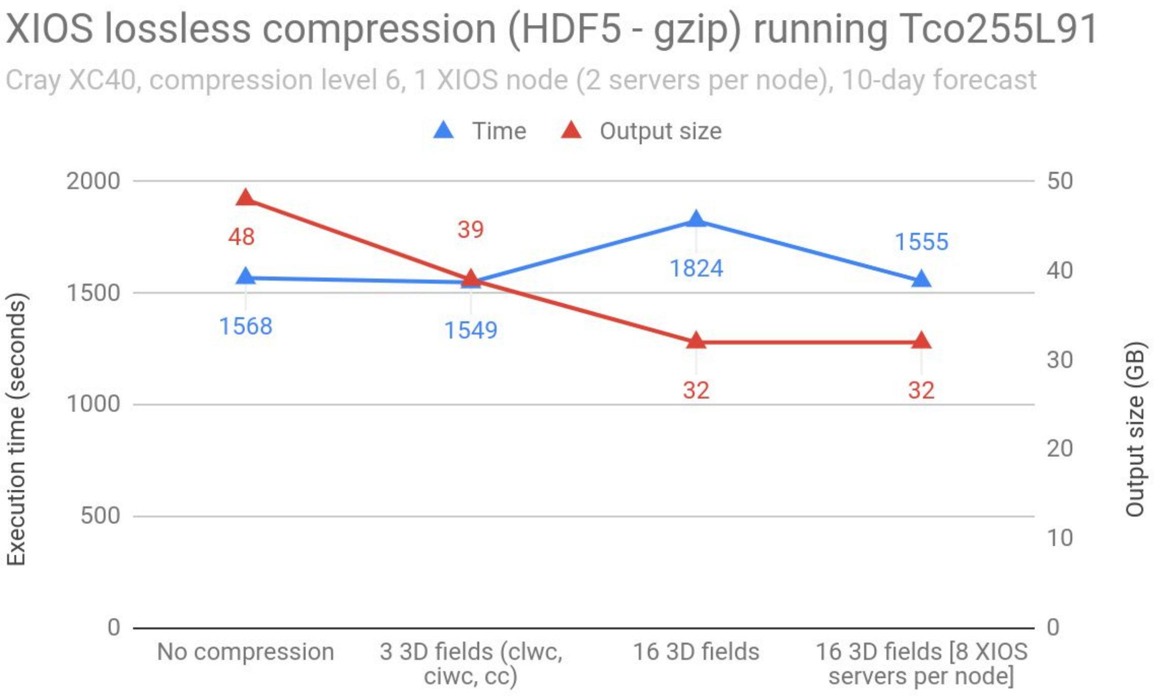
Lossy compression

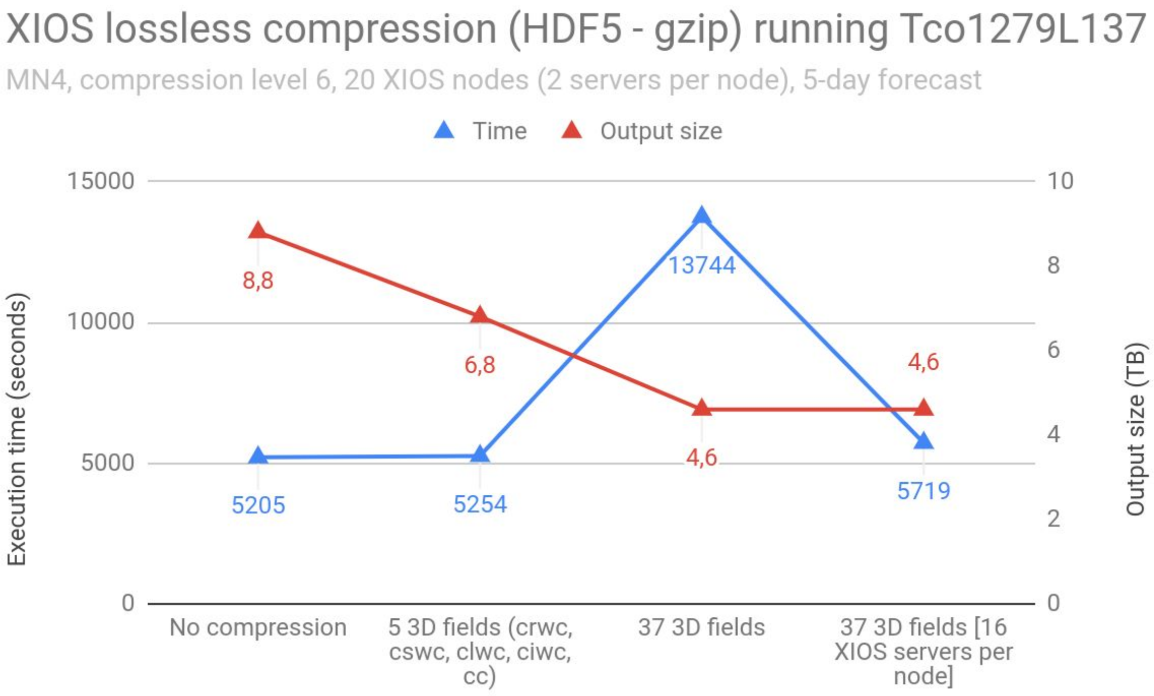
Barcelona
Supercomputing
Center
Centro Nacional de Supercomputación

What about XIOS compression?

- XIOS offers lossless data compression using gzip through HDF5.
- Although the overhead of outputting data through XIOS is really small for current resolutions, in the future this may well become a bottleneck because of the exponential growth of the output volume.
- The default lossless compression filter of HDF5 does not fit our needs:
  - If compression ratio is high, it takes too much time.
  - If it takes a reasonable amount of time, compression ratio is not enough.

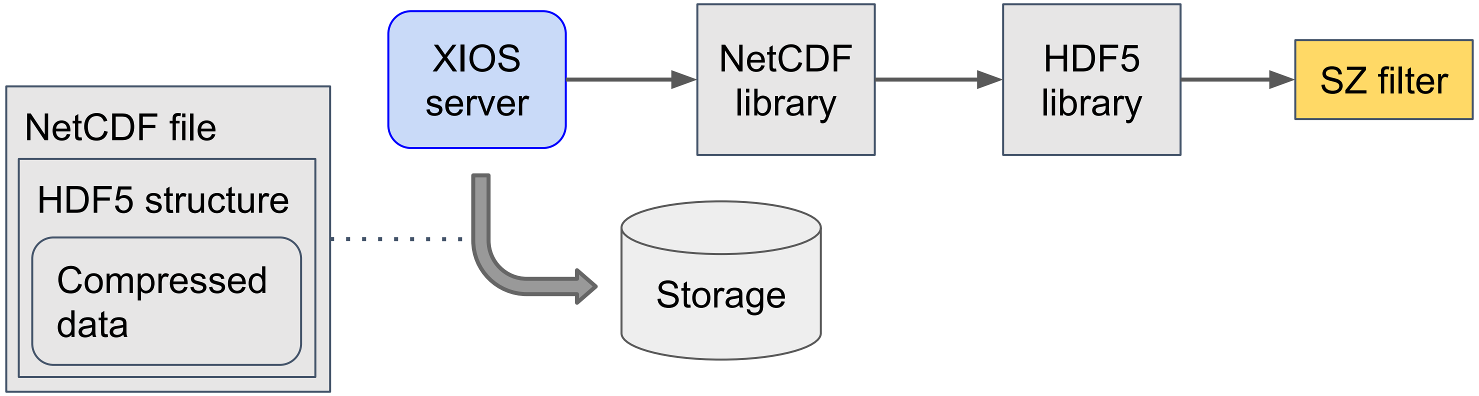
What about XIOS compression?



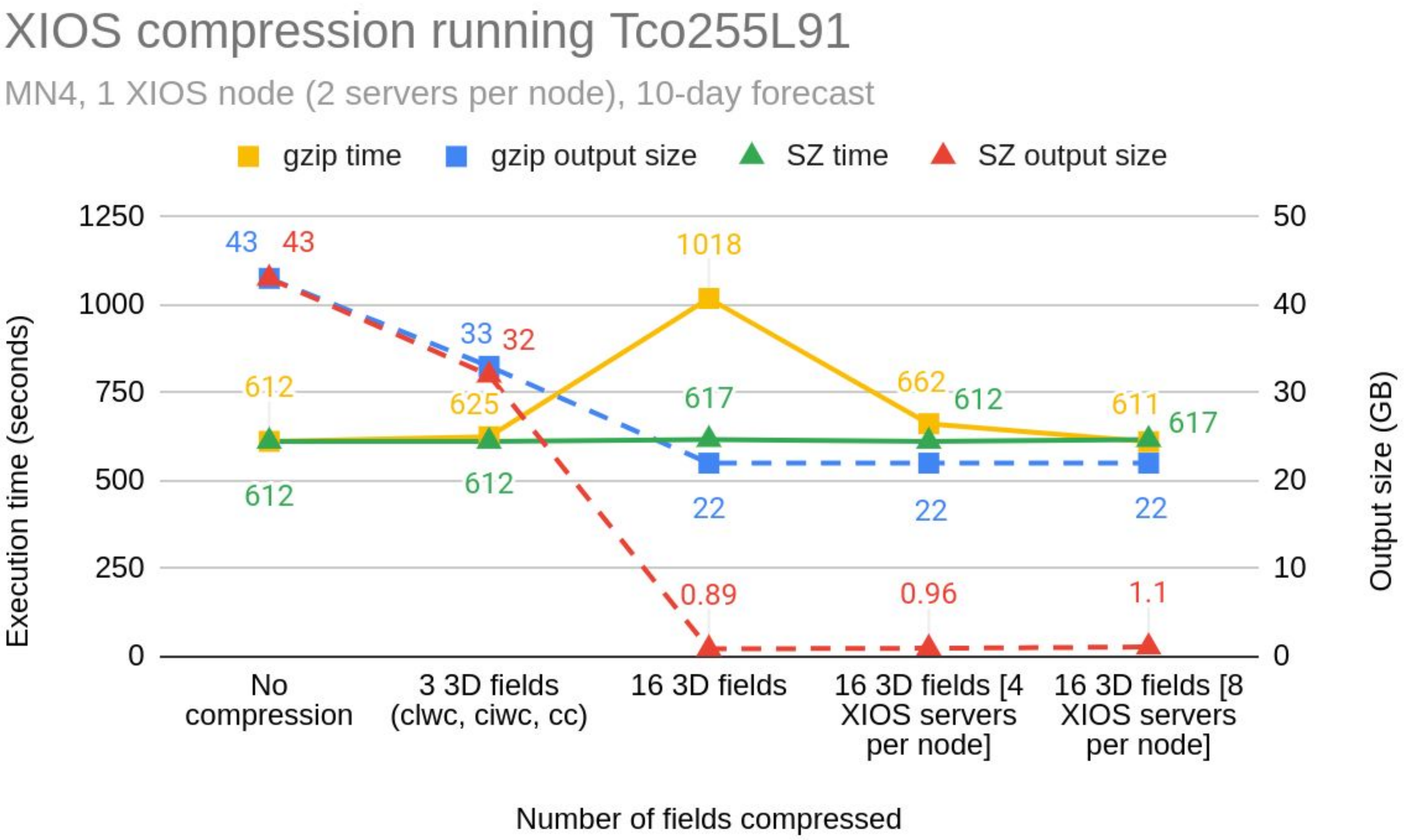


SZ lossy compression filter in XIOS

- Alternatively to lossless compression we wanted to test lossy compression and in particular, the SZ compressor from the Argonne National Laboratory (ANL).
- We started a collaboration with ANL to explore if SZ is suitable for XIOS regarding these points:
  - Reach high compression ratios.
  - Enough compression speed to considerably mitigate the I/O overhead.
  - Keep high accuracy.
- The SZ compressor is registered as a third-party filter of HDF5 which facilitates the integration in XIOS:



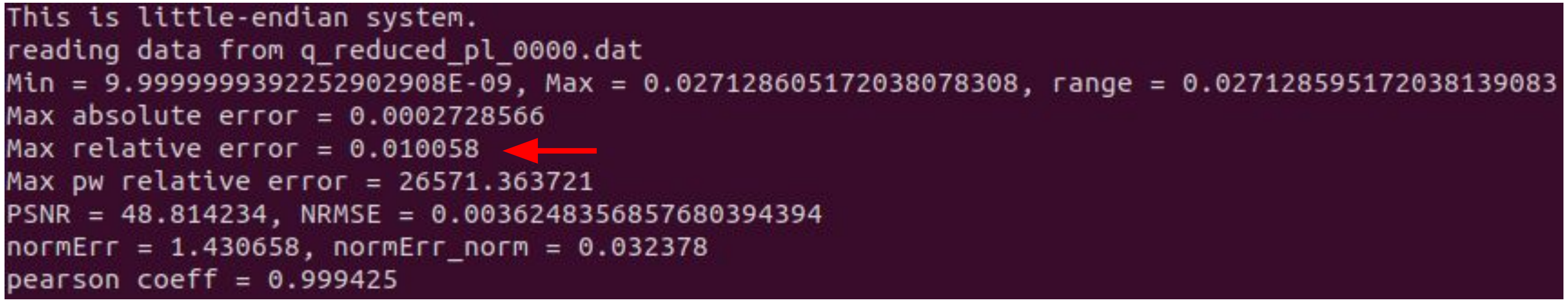
gzip vs. SZ compression: preliminary results



- gzip: compression level 6
- SZ: relative error bound 0.01

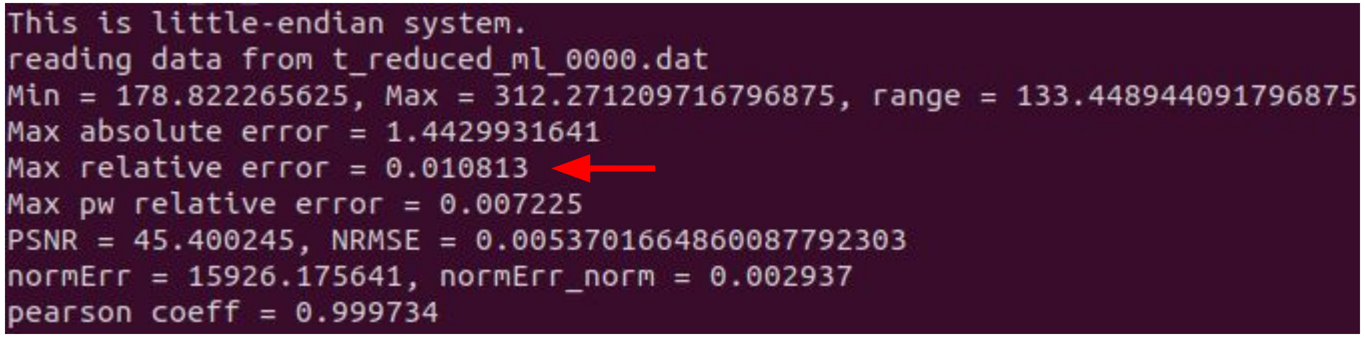
Technical validation

- Specific humidity (q):



```
This is little-endian system.
reading data from q_reduced_pl_0000.dat
Min = 9.9999999392252902908E-09, Max = 0.027128605172038078308, range = 0.027128595172038139083
Max absolute error = 0.0002728566
Max relative error = 0.010058
Max pw relative error = 26571.363721
PSNR = 48.814234, NRMSE = 0.0036248356857680394394
normErr = 1.430658, normErr_norm = 0.032378
pearson coeff = 0.999425
```

- Temperature (t):



```
This is little-endian system.
reading data from t_reduced_ml_0000.dat
Min = 178.822265625, Max = 312.271209716796875, range = 133.448944091796875
Max absolute error = 1.4429931641
Max relative error = 0.010813
Max pw relative error = 0.007225
PSNR = 45.400245, NRMSE = 0.0053701664860087792303
normErr = 15926.175641, normErr_norm = 0.002937
pearson coeff = 0.999734
```

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

esiwace
CENTRE OF EXCELLENCE IN SIMULATION OF WEATHER AND CLIMATE IN EUROPE

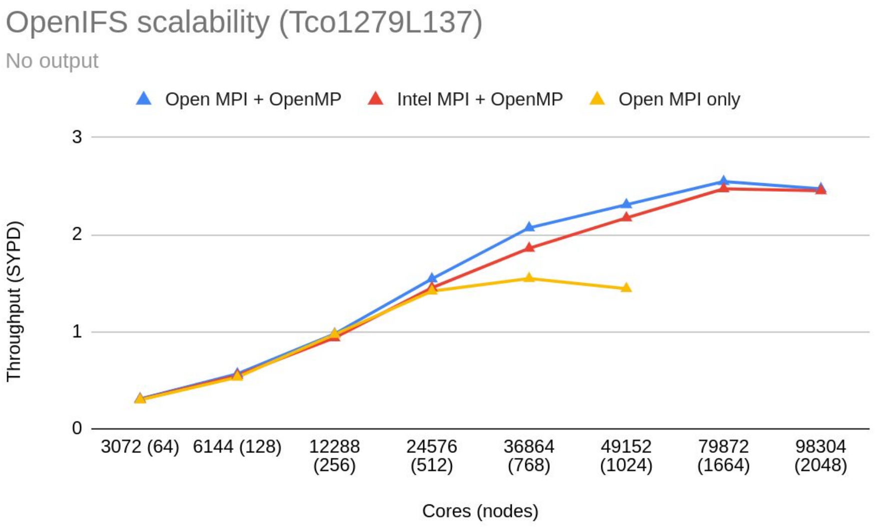
Scalability tests

Barcelona
Supercomputing
Center
Centro Nacional de Supercomputación

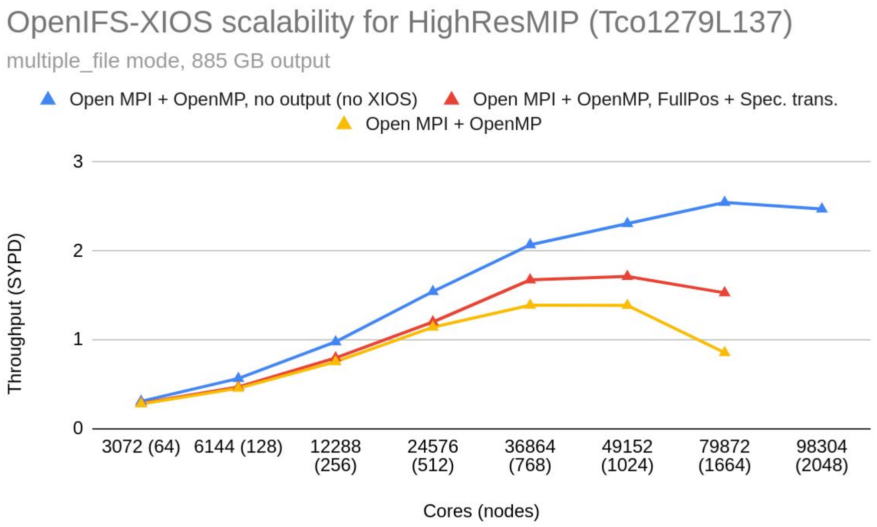
Benchmarking

- We consider it is important to anticipate the computational behaviour of OpenIFS for new pre-exascale machines, such as the upcoming MareNostrum 5 (MN5) of the BSC.
- OpenIFS is therefore benchmarked on the MareNostrum 4 (MN4) petascale machine to detect potential computational issues from:
    - Developments of new features.
    - Investigate if previous known limitations are solved.
- The benchmarking consists of large strong scaling tests by running different output configurations, such as changing multiple XIOS parameters and number of 2D and 3D fields. These tests use tens of thousands of cores, more than the 60% of MN4.
- A 9 km global horizontal resolution (Tco1279) is used as well as three different I/O configurations: no output, CMIP6-based fields and huge output volume (8.8 TB) to stress the I/O part.

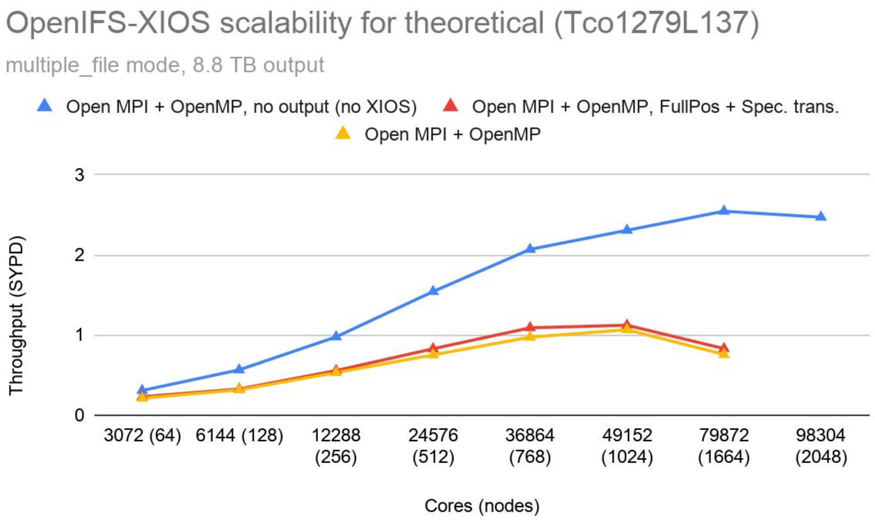
Scalability results


OpenIFS scalability (Tco1279L137)

No output


OpenIFS-XIOS scalability for HighResMIP (Tco1279L137)

multiple_file mode, 885 GB output


OpenIFS-XIOS scalability for theoretical (Tco1279L137)

multiple_file mode, 8.8 TB output

Scalability results

- OpenIFS scales up to 1664 nodes, but it should be more efficient.
- The MPI+OpenMP hybridisation is more efficient than ‘MPI only’ starting at 512 nodes.
- Open MPI is slightly better than Intel MPI.
- Production throughput depends on different factors such as I/O frequency and size, but it is strongly linked to the FullPos post-processing as well as the spectral transformations. It is necessary to further investigate the weight of each one.
- The XIOS scalability has to be improved, especially the memory management.

Conclusion

Barcelona
Supercomputing
Center
Centro Nacional de Supercomputación

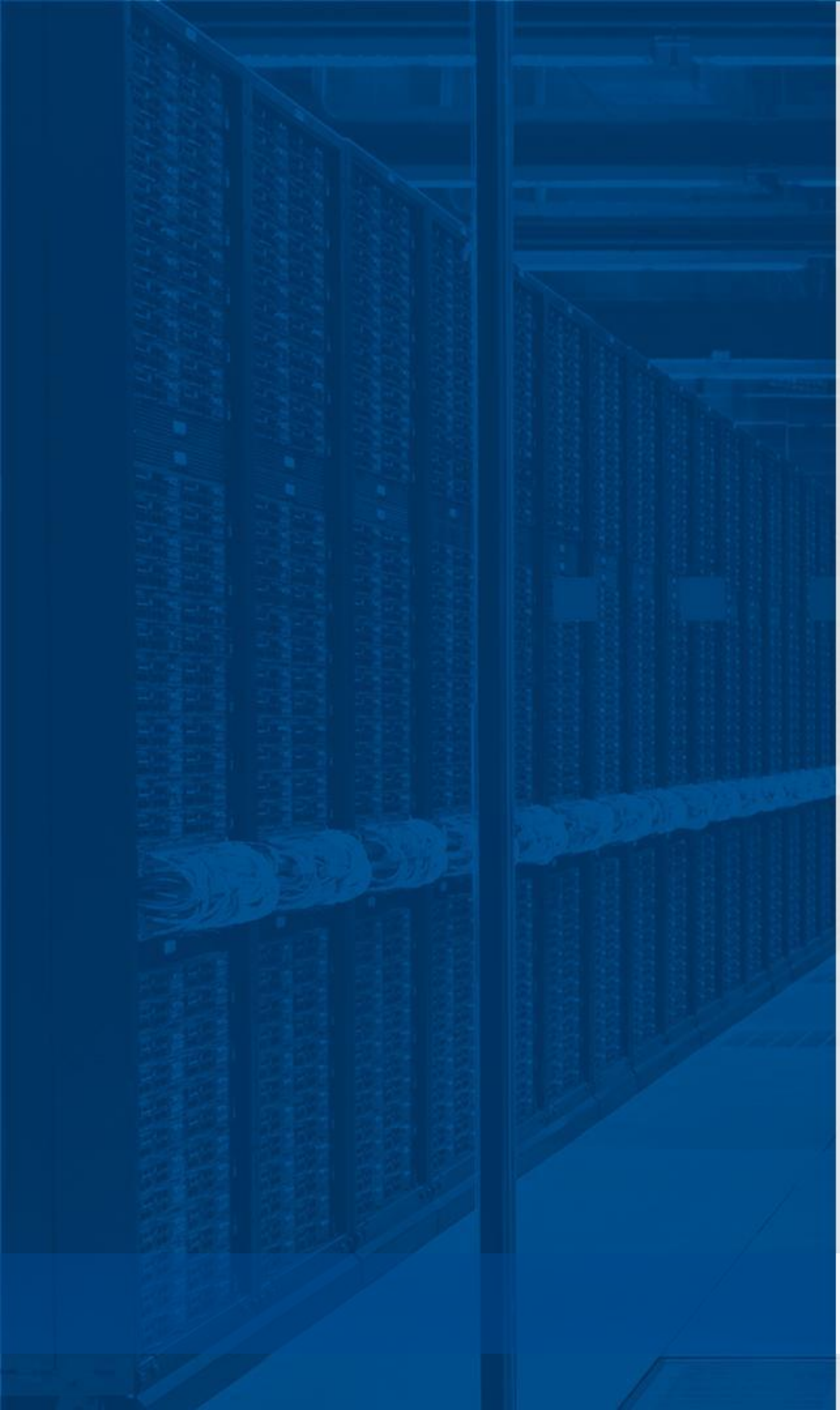
Conclusion

The computational efforts presented contribute to approach OpenIFS to the new upcoming HPC landscape:

- The XIOS server outperforms the sequential I/O scheme.
- In addition, XIOS provides OpenIFS with a more flexible output tool to accelerate costful post-processing tasks.
- The SZ lossy compressor is faster than the default gzip lossless compressor, achieving much higher compression ratios.
- Benchmarks suggest that OpenIFS scales up to 1664 nodes when using the hybrid approach MPI+OpenMP, and up to 768 nodes when enabling output through XIOS. However, more computational performance work is necessary to continue improving the efficiency.

Ongoing and future work

- Benchmarking and profiling of XIOS:
  - Scalability on pre-exascale machines.
  - Memory management.
- Continue working on lossy compression:
  - Discuss with climate scientists the adequate compression parameters (error bounds) for each field depending on the acceptable errors.
  - Test the SZ filter with the experimental HDF5 parallel I/O.
- Explore the benefits of applying OpenMP tasks (taskification) and Dynamic Load Balancing (DLB) to IFS.



Barcelona
Supercomputing
Center
Centro Nacional de Supercomputación

Thank you

esiwace
CENTRE OF EXCELLENCE IN SIMULATION OF WEATHER
AND CLIMATE IN EUROPE

xavier.yepes@bsc.es